MOLECULAR AND CELLULAR BIOLOGY, June 1986, p. 2125–2136
0270-7306/86/062125-12$02.00/0

Vol. 6, No. 6

Upstream Regions of the Human Cardiac Actin Gene That Modulate Its Transcription in Muscle Cells: Presence of an Evolutionarily Conserved Repeated Motif

ADRIAN MINTY† AND LARRY KEDES*

The MEDIGEN Project, Department of Medicine, Stanford University School of Medicine and Veterans Administration Medical Center, Palo Alto California 94305

Received 10 December 1985/Accepted 28 February 1986

Transfection into cultured cell lines was used to investigate the transcriptional regulation of the human cardiac actin gene. We first demonstrated that in both human heart and human skeletal muscle, cardiac actin mRNAs initiate at the identical site and contain the same first exon, which is separated from the first coding exon by an intron of 700 base pairs. A region of 485 base pairs upstream from the transcription initiation site of the human cardiac actin gene directs high-level transient expression of the bacterial chloramphenicol acetyltransferase gene in differentiated myotubes of the mouse C2C12 muscle cell line, but not in mouse L fibroblast or rat PC-G2 pheochromocytoma cells. Deletion analysis of this region showed that at least two physically separated sequence elements are involved, a distal one starting between −443 and −395 and a proximal one starting between −177 and −118, and suggested that these sequences interact with positively acting transcriptional factors in muscle cells. When these two sequence elements are inserted separately upstream of a heterologous (simian virus 40) promoter, they do not affect transcription but do give a small (four- to fivefold) stimulation when tested together. Overall, these regulatory regions upstream of the cap site of the human cardiac actin gene show remarkably high sequence conservation with the equivalent regions of the mouse and chick genes. Furthermore, there is an evolutionarily conserved repeated motif that may be important in the transcriptional regulation of actin and other contractile protein genes.

Actin constitutes one of the major cellular components both in the cytoskeleton and the muscle sarcomere, and has been shown to exist in multiple isoforms which are expressed in a tissue-specific way (67). A pair of actin genes is expressed in the striated muscles (skeletal muscle and heart) of vertebrates. One of these genes (α-skeletal) encodes the major actin species in adult skeletal muscle actin, and the other gene (α-cardiac) encodes the major actin species in the adult heart (67). The cardiac actin gene has also been shown to be expressed at high levels in embryonic and fetal skeletal muscle (45, 49, 52, 61; A. J. Minty, E. Hardeman, P. Gunning, W. Bains, H. Blau, and L. H. Kedes, UCLA Symp. Mol. Cell. Biol., in press). In adult skeletal muscle, however, there is only a low level of cardiac actin mRNA, showing that the cardiac actin gene is subject to developmental regulation (28, 49). The different actin genes are unlinked in mammalian genomes (13, 29, 50) and are also unlinked to other contractile protein genes (14, 63). It thus appears that *cis*-acting sequences in the vicinity of each individual actin gene must be implicated in their differential regulation.

In a first attempt to characterize sequences important in the high-level, tissue, and developmental stage-specific transcription of the cardiac actin gene, we addressed the following questions. (i) Where is the transcription start site for the human cardiac actin gene? (ii) Is this transcription start site used both in heart and developing skeletal muscle? (iii) Are there sequence elements upstream of the transcription start site that can be implicated in tissue-specific transcriptional regulation? (iv) Are multiple sequence elements involved? (v) How do these sequences act to regulate transcription?

The bacterial chloramphenicol acetyltransferase (CAT) gene (24) has been used as a marker gene to demonstrate the presence of upstream sequences specifying tissue-specific regulation of genes expressed in pancreas (68), liver (11, 60), lens (10), and muscle (46). In this study we used the CAT system to characterize sequences upstream of the transcription start site of the human cardiac actin gene which are implicated in the high-level transcription of this gene in muscle cells. For these experiments, we used the C2C12 subclone (4) of the mouse C2 myoblast cell line isolated by Yaffe and Saxel (74). These cells can be induced to differentiate into multinucleate myotubes, accumulating very high levels of cardiac actin mRNA (equivalent to that in the adult mouse heart) in a period of 24 to 36 hours (1). In addition, since they are readily transfected by the calcium phosphate precipitation method (65) (see below), they represent a useful system for transient expression studies on the cardiac actin gene.

Our experiments describe two separate sequence elements in the highly conserved upstream segments of the human cardiac actin gene that modulate the tissue-specific transcription of this gene. Deletion of these sequences lowers transcription in muscle cells but has no effect on transcription in L cells, suggesting that these sequences interact with a tissue-specific transcriptional factor(s) in muscle cells. The distal upstream region contains sequences similar to viral and cellular enhancer elements. The sequence of the proximal region contains four evolutionarily conserved regions that share a common sequence pattern which may be involved in the transcriptional regulation of actin and other contractile protein genes.

* Corresponding author.
† Present address: Institut de Pathologie Moleculaire, 75674 Paris Cedex 14, France.

MATERIALS AND METHODS

Cells and cell culture. C2 cells were originally isolated by Yaffe and Saxel (74) by the selective serial passaging of myoblasts isolated from crushed thigh muscle of a 2-month-old C3H mouse. A well-differentiating subclone of these cells (C2C12 [4]) was obtained from Helen Blau. Mouse L fibroblast cells deficient in thymidine kinase (Ltk^-) were obtained from the laboratory of Stanley Cohen. The PC-G2 clonal cell line, derived from an experimentally induced rat pheochromocytoma tumor (23), was obtained from the laboratory of Eric Shooter. C2C12 cells were cultured as described previously (4). For differentiation into multinucleate myotubes, confluent cultures were switched to nutrient-poor medium (Dulbecco modified Eagle medium supplemented with 2% horse serum). Extensive myoblast fusion was obtained after 24 to 36 h (1). L cells and PC-G2 cells were maintained in Dulbecco modified Eagle medium supplemented with 10% fetal calf serum.

Construction of plasmids; preparation and sequencing of plasmid DNA. In general, the procedures suggested by Maniatis et al. (42) were followed for construction of plasmids and preparation of plasmid DNA. Restriction endonucleases, T4 DNA ligase, nuclease BAL 31, and phosphorylated *Eco*RI and *Hin*dIII synthetic linkers were purchased from New England Biolabs, Inc., Beverly, Mass. DNA polymerase I (Klenow fragment) was purchased from Bethesda Research Laboratories, Inc., Gaithersburg, Md. The 5′ deletions in the human cardiac actin gene were generated from the *Bam*HI site 0.8 kilobases (kb) upstream from the cap site (Fig. 1) by BAL 31 digestion, followed by filling the ends with DNA polymerase and adding *Eco*RI linkers. After the resulting DNA fragments were cut with *Eco*RI and *Bgl*II and gel purified, they were cloned in the plasmid pSVneo.HCA (51) cut with *Eco*RI and *Bgl*II. Plasmid minipreparations were digested with restriction endonucleases to determine the approximate size of the 5′-flanking sequence. Appropriate clones were cut with *Fnu*DII (which cuts at 68; Fig. 1). *Hin*dIII linkers were added at this site, and the DNA was cut with *Eco*RI. The *Eco*RI-*Hin*dIII fragments were recloned in a vector, pHCA0CAT, which contains the CAT gene *Hin*dIII-*Bam*HI fragment from pSV2CAT (24) cloned between the *Hin*dIII and *Bam*HI sites of pBR322 (Fig. 1D). The resulting series of plasmids were named pHCA485CAT to pHCA47CAT. The numbers following the HCA designation indicate the numbers of nucleotides 5′ to the cap site, determined by sequencing the *Eco*RI-*Hin*dIII fragments in M13.mp11 (48).

Large-scale preparations of plasmid DNA were prepared from cleared bacterial lysates by banding on ethidium bromide-cesium chloride gradients. All preparations to be used for quantitative CAT assays were purified on two successive gradients. Nucleotide sequencing was performed by both the chemical degradation method (44) and the dideoxy chain terminator method (65) with the M13 vectors mp11 and mp10 (48).

Cell transfection and CAT assays. L and PC-G2 cells were split 1:10 into 100-mm-diameter dishes the day before transfection to give $\sim 10^6$ cells per dish for transfection. For C2C12 cells, confluent 100-mm-diameter dishes were split 1:4 the day before to give a semiconfluent dish for transfection. Transfection mixtures were prepared as follows (per 100-mm-diameter dish). To 10 μg of DNA in 0.5 ml of 2× *N*-2-hydroxyethylpiperazine-*N*′-2-ethanesulfonic acid (HEPES)-buffered saline (42 mM HEPES [pH 7.05], 270 mM NaCl, 10 mM KCl, 1.4 mM Na_2HPO_4, 11 mM dextrose) was added dropwise 0.5 ml of a solution of 250 mM $CaCl_2$. This was done by constant mixing obtained by bubbling a stream of air through the DNA solution. The calcium phosphate-DNA precipitate was left for 20 to 30 min at room temperature and then was added to a 100-mm-diameter dish of cells from which medium had been removed and left for 20 min over the cells at room temperature. A total of 10 ml of medium (10% fetal calf serum in Dulbecco modified Eagle medium) was added, and the cells were left for 8 to 20 h at 37°C. A glycerol shock step (40) was not included because it was found that under these transfection conditions, this step made no difference in the resulting CAT activity. The medium was then changed to the normal growth medium for these cells, or fusion medium in the case of C2C12 cells, and the cells were left for an additional 24 to 48 h. For CAT assays, cell extracts were prepared by three cycles of freezing and thawing in 100 μl of 250 mM Tris hydrochloride (pH 7.8), and 10 to 50 μl of extract was incubated with 0.1 to 0.2 μCi of [^{14}C]chloramphenicol (New England Nuclear Corp., Boston, Mass.) and analyzed on silica gel thin-layer chromatography plates as described by Gorman et al. (24). Quantitation of CAT assays was performed by scintillation counting of the appropriate areas of the chromatogram.

RNA preparation and S1 nuclease analysis. RNA from 4-day-postfusion human myotube cultures and from adult human heart was a gift from P. Gunning and E. Hardeman and was prepared as described by Gunning et al. (28). Single-stranded probes corresponding to the first exon of the human cardiac actin gene were prepared from M13.mp10 clones containing an *Eco*RI-*Hin*dIII insert representing nucleotides −47 to 68 of the gene (Fig. 1C). Approximately 5 ng of M13 sequencing primer annealed to approximately 500 ng of single-stranded M13 DNA in 25 μl of 20 mM Tris hydrochloride (pH 8.5)–10 mM $MgCl_2$ was extended by adding 6 U of DNA polymerase 1 (Klenow fragment) in 40 μl of 5 mM dithiothreitol–5 mM Tris hydrochloride (pH 8.5)–250 μM dATP–250 μM dGTP–100 pmol of [α-^{32}P]dCTP–100 pmol of [α-^{32}P]dTTP (>400 Ci/mmol; Amersham Corp., Arlington Heights, Ill.) and by incubating for 1 h at 30°C. The reaction products were restricted with *Eco*RI, and a single-stranded probe complementary to the insert was isolated on a 6% acrylamide-urea sequencing gel. Approximately 3×10^4 cpm of this probe was hybridized with 5 to 20 μg of RNA in 20 μl of 80% formamide–10 mM piperazine-*N*-*N*′-bis(2-ethanesulfonic acid) (PIPES; pH 6.4)–1 mM EDTA. After denaturation at 80°C for 3 min, hybridization was carried out at 48°C for 16 h. S1 nuclease digestion was performed by dilution into 200 μl of 260 mM NaCl–30 mM sodium acetate (pH 4.6)–1 mM $ZnCl_2$ or $ZnSO_4$–20 μg of denatured calf thymus DNA per ml, using 250 to 500 U of S1 nuclease (Sigma Chemical Co., St. Louis, Mo.), for 5 to 20 min at 37°C. Carrier tRNA was added, and undigested nucleic acids were precipitated with isopropanol and analyzed on 8% acrylamide-urea sequencing gels.

RESULTS

Characterization of the first exon of the human cardiac actin gene. Results of previous experiments have indicated that the cardiac actin gene is present as a single copy in both the mouse (49) and human (28) genomes and is expressed both in heart and skeletal muscle. However the transcription initiation site for a mammalian cardiac actin gene has not been characterized previously. The 6-kb human cardiac actin gene is contained within a 13-kb *Eco*RI fragment (20, 30). A map of this fragment showing the location of strategic restriction endonuclease sites is shown in Fig. 1A. About 800 nucleo-

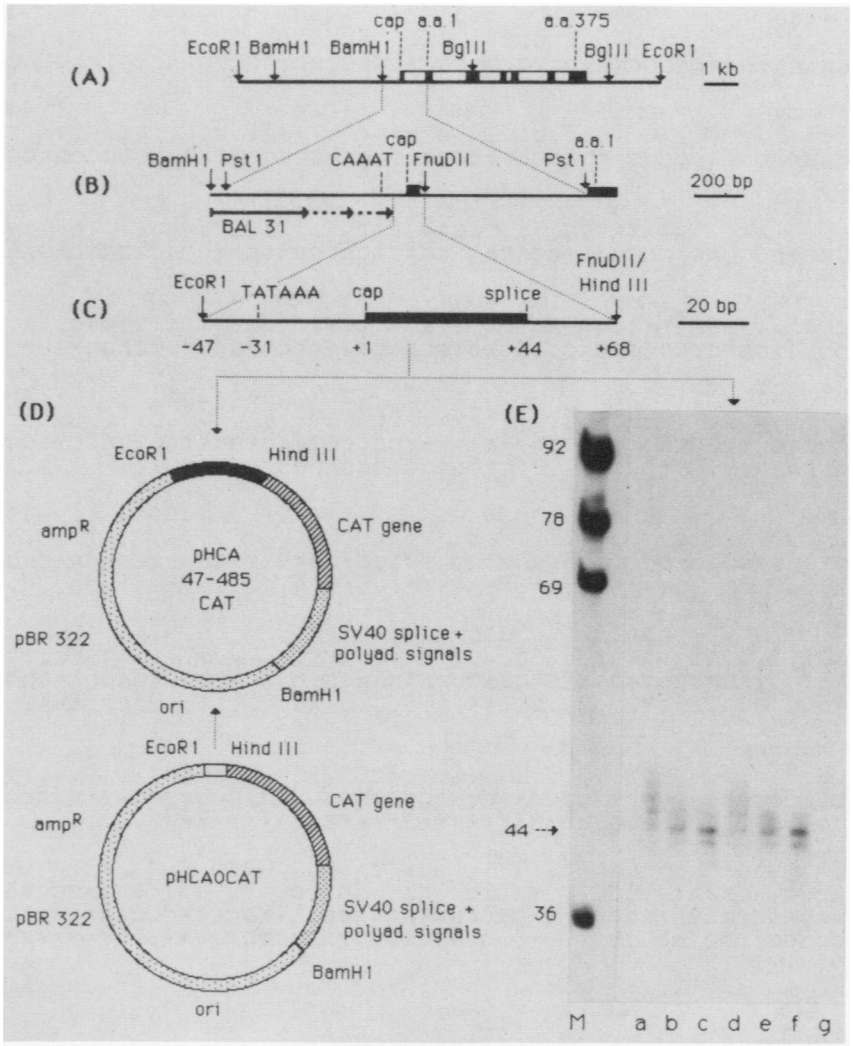

FIG. 1. Characterization and cloning of fragments from the 5′ end of the human cardiac actin gene. (A) Map of pHRL83 (20) showing strategic restriction sites. Exons are indicated by solid boxes, and introns are indicated by open boxes. The scales of each of the three maps are different and are indicated by the horizontal bars. a.a., Amino acid. (B) The upstream region of the human cardiac actin gene from the *Bam*HI site 800 bp upstream of the transcription start (cap) site to the translation initiation site (amino acid 1). The position of the sequence CAAAT discussed in the text is indicated. (C) Creation of 5′ deletions. Fragments containing different lengths of upstream sequence were generated by BAL 31 digestion from the upstream *Bam*HI site. *Eco*RI linkers were added at their 5′ ends, the fragments were recut with *Fnu*DII (at 68), and *Hin*dIII linkers were added at their 3′ ends. The shortest *Eco*RI-*Hin*dIII fragment containing 47 bp of upstream sequence is shown, and the position of the sequence TATAAA discussed in the text is indicated. (D) Maps of plasmids used for construction of human cardiac actin-CAT gene vectors. The *Hin*dIII-*Bam*HI fragment from pSV2CAT (24) containing the bacterial CAT gene with the SV40 small t intron and early-region polyadenylation (polyad.) signals was inserted in pBR322 to generate pHCA0CAT. *Eco*RI-*Hin*dIII fragments from panel C were cloned in pHCA0CAT to generate the series of plasmids pHCA485CAT to pHCA47CAT. The circular plasmid molecules are not drawn to scale. (E) Determination of the transcription start site of the gene. The *Eco*RI-*Hin*dIII fragment shown in panel C was cloned in mp10; and a single-stranded probe, prepared as described in the text, was hybridized to RNA from human myotube cultures (5 μg) and adult human heart (20 μg). The resulting hybird molecules were treated with S1 nuclease for different lengths of time and then analyzed on a denaturing 8% acrylamide gel. Lane a, myotube RNA, 5 min of S1 nuclease digestion; lane b, myotube RNA, 10 min of S1 nuclease digestion; lane c, myotube RNA, 20 min of S1 nuclease digestion; lane d, heart RNA, 5 min of S1 nuclease digestion; lane e, heart RNA, 10 min of S1 nuclease digestion; lane f, heart RNA, 20 min of S1 nuclease digestion; lane g, tRNA, 20 min of S1 nuclease digestion. The size markers (M) were pBR322 cut with *Hpa*II and labeled with T4 DNA polymerase, as described in the accompanying paper (51). The identification of the major protected fragment of 44 nucleotides is indicated. This result, in conjunction with the identification of the likely location of the 5′-splice site of the first intron, allows assignment of the transcription initiation site.

tides upstream of the translation start site is a consensus TATAAA signal that is characteristic of the transcription start sites of many eucaryotic genes (Fig. 2). To prove that this region contains the transcription initiation site, we hybridized RNA from human skeletal muscle cultures to a uniformly labeled single-stranded DNA probe corresponding to a fragment from 16 nucleotides upstream of this TATAAA box to 99 nucleotides downstream. The autoradiogram shown in Fig. 1E (lane c) demonstrates that the RNA protected fragments of 43 to 44 nucleotides from digestion by S1 nuclease. This result demonstrates that the first exon of the human cardiac actin gene is 43 to 44 base pairs (bp) long and that the first intron is located in the 5′-untranslated region of the gene. The location of the 5′ splice (donor)

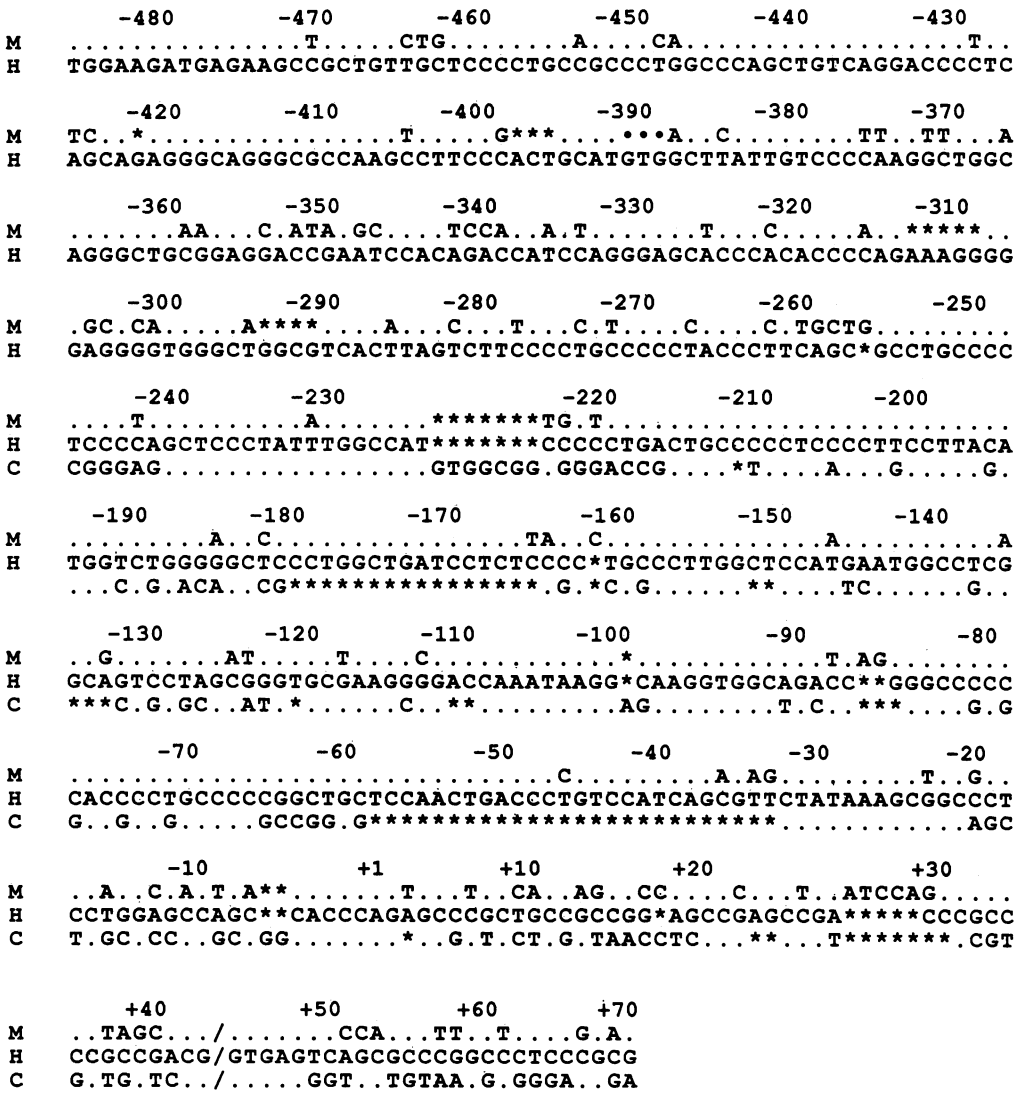

FIG. 2. Evolutionary conservation of the upstream regions of vertebrate cardiac actin genes. The nucleotide sequence of the upstream region of the human cardiac actin gene was determined both from dideoxy sequencing of the *Eco*RI-*Hin*dIII fragments resulting from the BAL 31 deletions (Fig. 1) cloned in M13mp11 or M13mp10 and from sequencing of the BAL 31 deletions and from the *Msp*I (−87), *Ava*II (−112), and *Hin*fI (47, −349) sites, as described by Maxam and Gilbert (44). This human (H) sequence was compared with the corresponding regions of the mouse (M) (6; Garner and Buckingham, personal communication) and chick (C) (8) cardiac actin genes. Bases identical to the human sequence are designated with a dot, and the differences are shown. Bases missing in one of the cardiac actin gene sequences, when compared with one of the other sequences, are designated by an asterisk. The location of a sequence $(TG)_{24}$ in the mouse sequence is designated by three closed circles. The location of the 5′ splice site of the first intron is indicated by the slash at base 44.

junction is unknown, but within the probe DNA there is a presumptive splice junction, CG/GTGAGT, that corresponds to the splice junction of the chick cardiac actin gene transcript (8). If this sequence is also the splice junction for the human transcript, then, in conjunction with the result of the S1 nuclease protection data, we propose that the transcriptional initiation (cap) site is the adenine residue 44 nucleotides upstream from this splice site and 31 nucleotides downstream from the first thymine of the TATAAA box. This adenine residue is labeled 1 in Fig. 2.

When the probe was hybridized to RNA from adult human heart, fragments of identical sizes were observed after S1 nuclease digestion (Fig. 1E, lanes a through c and d through f), showing that this exon is expressed in both skeletal muscle and heart. The amounts of this 5′ probe protected by the heart and myotube samples corresponded to the relative abundance of cardiac actin mRNA in these two samples (1:4) which was determined previously with a 3′-noncoding probe specific for cardiac actin mRNA (28; Minty et al., in press; E. Hardeman, R. Wade, and P. Gunning, unpublished data). Thus, we conclude that the major initiation site for cardiac actin transcription in the two striated muscles is the same.

The upstream regions of vertebrate cardiac actin genes show striking conservation. Some 82% of the bases of the first exon of the gene and the adjacent 80 bp of intron sequence are guanine or cytosine, and this region contains 22 CpG doublets (Fig. 2; unpublished data). This region thus constitutes a CpG-rich island because the surrounding DNA contains very few of these dinucleotides. Such CpG-rich islands have been suggested to be preferentially associated with housekeeping genes and, possibly, to be involved in the continuous activity of these genes (3, 72). In the case of the human

cardiac actin gene, this gene is not continuously expressed, and the CpG-rich region does not extend into the region which we implicate (see below) in transcriptional regulation in muscle cells. Its function therefore is not clear. In addition, the CpG doublets show a very low degree of evolutionary conservation; only 3 of the 11 CpGs between −40 and 50 in the human gene are present in the mouse gene (Fig. 2).

In contrast, the nucleotide sequence of the region upstream of the TATAAA sequence in the human gene shows a striking degree of similarity with the corresponding regions of the mouse (6; I. Garner, personal communication) and chick (8) cardiac actin genes (Fig. 2). The region from the TATAAA sequence to −240 is 90% conserved between the human and mouse genes and ~70% conserved between the human and chick genes. The human and mouse genes also show considerable similarity (approximately 77%) over the region from −240 to −485, as far as our sequence comparison extends. This high degree of similarity has also been reported for the 3′-untranslated regions of these genes (8, 27, 45). It is important to note, however, that other regions of these genes such as the 5′-untranslated regions show a lower degree of similarity (65% between the human and mouse genes).

Tissue-specific transcription from an upstream region of the human cardiac actin gene. We tested the hypothesis that the evolutionary constraint on the upstream sequences of cardiac actin genes may be due to a role in tissue-specific transcriptional regulation. We constructed a plasmid (pHCA485CAT) containing 485 bp of human cardiac actin upstream sequence plus the first exon and 24 bp of the first intron linked at the *Fnu*DII site at 68 to the bacterial CAT marker gene (24) which carries a simian virus 40 (SV40) intron and polyadenylation site, to ensure proper processing of transcripts. Transfection of this plasmid into myoblasts of the mouse C2C12 muscle cell line, by the calcium phosphate precipitation method (25), led to high levels of transient CAT expression when these cells were allowed to differentiate into myotubes (Fig. 3, lanes e and f). In contrast, transfection of pHCA485CAT into the rat PC-G2 pheochromocytoma cell line (Fig. 3, lanes b and c) led to only low-level CAT expression. As a control for differences in transfection efficiencies and overall transcription levels in the two cell lines, we also transfected these cells with the plasmid pSV2CAT (24), which contains the SV40 early promoter. This plasmid gave a similar CAT signal in the two cell lines (Fig. 3, lanes a and d). In this particular experiment pSV2CAT gave a fourfold higher signal than did pHCA485CAT in C2C12 cells. However, in subsequent experiments with plasmid DNA preparations purified through a second ethidium bromide-cesium chloride gradient (Fig. 4), we observed comparable (50 to 100%) CAT signals with the two plasmids in the C2C12 muscle cell line. The level of CAT expression from pHCA485CAT was 20- to 40-fold greater in the C2C12 muscle cell line than in the PC-G2 pheochromocytoma cell line (compare lane f with lane c in Fig. 3). A second nonmuscle cell, the mouse fibroblastic L-cell line, also showed a low level of CAT expression from pHCA485CAT, which was similar to the signal seen in PC-G2 cells (Fig. 5, lane b; see below). These results indicate that the upstream region of the cardiac actin gene contains a sequence that enables high-level transcription specifically in muscle cells.

Two explanations for these results are the interaction of this region with *trans*-acting positive regulatory factors in muscle cells (17, 47, 68) or its interaction with negative regulatory factors in nonmuscle cells (5, 37). To distinguish between these possibilities we constructed a series of vectors containing decreasing lengths of 5′-flanking sequence (pHCA443CAT-pHCA47CAT) with BAL 31 nuclease. When these vectors were transfected into C2C12 muscle cells or L fibroblast cells, it was seen that the deletions considerably decreased transcriptional activity in the C2C12 cells (40- to

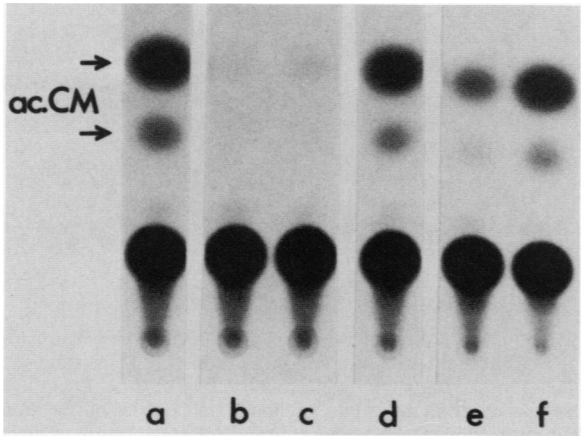

FIG. 3. Transcriptional activity of the upstream region of the human cardiac actin gene in rat PC-G2 cells and C2C12 myotubes. Plasmid DNA (10 μg) was introduced into proliferating C2C12 myoblast cells and rat PC-G2 pheochromocytoma cells by calcium phosphate precipitation. The C2C12 cells were allowed to fuse into multinucleated myotubes. Transient levels of CAT expression were determined as described in the text with 10 or 40 μl of cell extract. The 1- and 3-mono-acetylated chloramphenicol (ac-CM) forms were separated from unacetylated chloramphenicol by thin-layer chromatography (24). For each of the assays, the cell type, plasmid, and volume of cell extract assayed are indicated, respectively: lane a, PCG2 cells, pSV2CAT, 10 μl; lane b, PCG2 cells, pHCA485CAT, 10 μl; lane c, PC-G2 cells, pHCA485CAT, 40 μl; lane d, C2C12 cells, pSV2CAT, 10 μl; lane e, C2C12 cells, pHCA485CAT, 10 μl; lane f, C2C12 cells, pHCA485CAT, 40 μl.

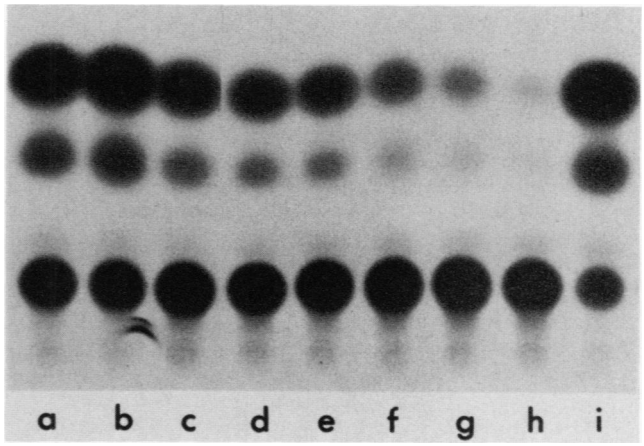

FIG. 4. Transcriptional activity of upstream regions of the human cardiac actin gene in C2C12 myotubes. Plasmid DNAs were transfected into C2C12 cells, and CAT activity from equal volumes of cell extract was determined as described in the legend to Fig. 3. Lane a, pHCA485CAT; lane b, pHCA443CAT; lane e, pHCA395CAT; lane d, pHCA263CAT; lane e, pHCA177CAT; lane f, pHCA118CAT; lane g, pHCA101CAT; lane h, pHCA47CAT; lane i, pSV2CAT.

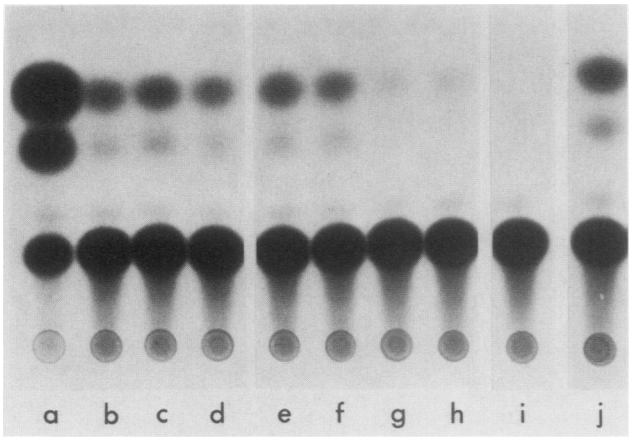

FIG. 5. Transcriptional activity of upstream regions of the human cardiac actin gene in L cells. Plasmid DNAs were transfected into mouse L cells, and CAT activity from equal volumes of cell extracts was measured as described in the legend to Fig. 3. Lane a, pSV2CAT; lane b, pHCA485CAT; lane c, pHCA177CAT; lane d, pHCA118CAT; lanes e and f, two preparations of pHCA47CAT; lanes g and h, two preparations of pHCA0CAT; lane i, no extract; lane j, pA10CAT2.

50-fold; for example, Fig. 4, lanes a and h) without affecting the low level of activity in L cells (Fig. 5, lanes b through f). Although results of these experiments do not exclude the possibility of repressor molecules in L cells, they strongly suggest the interaction of positively acting regulatory factors with this region in C2C12 cells.

At least two upstream regions are important in this transcriptional regulation. These deletion experiments also allowed us to begin to define the sequence elements involved in the interaction with regulatory factors. To overcome the variability inherent in this type of experiment (33), we performed these experiments several times (four times for most plasmid constructions, with a minimum of two times) using independent preparations of plasmid DNA. Results of these experiments indicate that the major losses in transcriptional activity in myotubes accompany the deletions −443 to −395 (from 100 to ~60%) and −177 to −47 (from ~40 to ~2%) (Fig. 6). Inclusion of a longer upstream sequence from the cardiac actin gene (up to the *Bam*HI site at ~800 bp) does not significantly alter the CAT signal obtained in C2C12 cells. Similarly, deletion of sequences between the two segments from −395 to −177 had little effect on the transcriptional activity of the cardiac actin promoter in these experiments. This indicates that two distinct regions (a distal region, D, beginning between −443 and −395, and a proximal region, P, beginning between −177 and −118) are important in the tissue-specific transcriptional regulation of the cardiac actin gene.

Within region P, deletion of nucleotides 118 to 101 removed the sequence CCAAAT which has been designated (8, 56, 58) as the putative CAAT box (2) of vertebrate striated muscle actin genes. Plasmid pHCA118CAT showed an approximately 5-fold higher transcriptional level in C2C12 cells than did pHCA47CAT (the values were 3.5-, 5-, 7-, and 7-fold in four different experiments), whereas pHCA101CAT showed only an average 2-fold higher level than pHCA47CAT (the values were 1.3-, 1.3-, 2.6-, and 3-fold; for example, Fig. 4, lanes f through h). In contrast these deletions had little effect on the CAT signal obtained in L cells (Fig. 4, lanes d, e, and f).

Although the levels of CAT activity observed in L cells with the vectors that carry segments of the human cardiac actin gene upstream region were low, they were three- to fourfold higher than those obtained with the plasmid pHCA0CAT (compare lanes b to f with lanes g and h in Fig. 5). pHCA0CAT contains the CAT gene inserted in pBR322 in the same location as the other pHCA constructs, but without cardiac actin upstream sequences. Thus, CAT expression in L cells of vectors carrying segments of the human cardiac actin gene upstream region appears significant. Furthermore, although the levels of CAT activity were low, they still represent the activity of a moderately strong promoter. For example, we have observed comparable CAT signals with the pA10CAT2 plasmid (39) (Fig. 5, lane j), which contains the SV40 early promoter, including its 21-bp repeats, but lacks the 72-bp regulatory elements (39). We

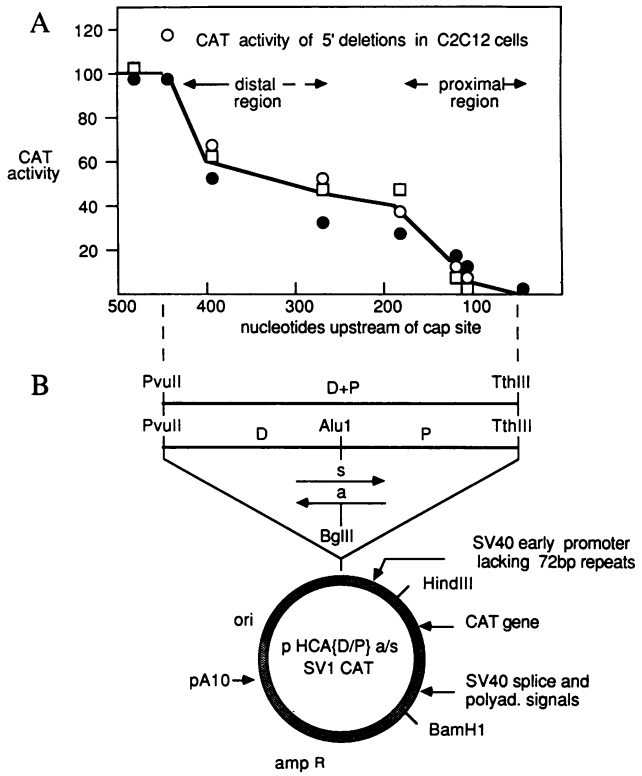

FIG. 6. Quantitation of the transcriptional acticity of upstream regions of the human cardiac actin gene in C2C12 myotubes and cloning of these regions in the plasmid pA10CAT2. (A) Quantitation of the transcriptional activity of upstream regions of the human cardiac actin gene in C2C12 myotubes. CAT assays were performed as described in the legend to Fig. 3 and quantitated by scintillation counting of the regions of the chromatogram containing the acetylated forms of chloramphenicol. CAT activities were normalized to the activity of pHCA485CAT in each experiment, which was taken as 100%. The quantitation of three experiments with two different sets of plasmid DNA preparations (○ and ●, and □) is shown. Each data point represents the average of two CAT assays on this cell extract. (B) Schematic representation of insertion into the *Bgl*II site of pA10CAT2 (39) of upstream regions of the human cardiac actin gene. The inserted fragments are the *Pvu*II-*Tth*III, *Pvu*II-*Alu*I, and *Alu*I-*Tth*III fragments representing regions D-P, D, and P, respectively, in both sense (s) and antisense (a) orientations relative to the SV40 promoter. The transcriptional activities of the resulting plasmids (designated pHCA[D-P, D, or P] (a or s).SV1CAT) are illustrated in Fig. 7. polyad., Polyadenylation.

conclude that sequences upstream from −47 are not required for the activity of the cardiac actin promoter in L cells.

To confirm this observation, we inserted the 72-bp regulatory element from SV40 (69) into the plasmids pHCA0CAT, pHCA47CAT, and pHCA177CAT at a position downstream from the CAT gene. The presence of the 72-bp regulatory element in these constructions increased the CAT expression from these plasmids in L cells 10- to 20-fold. The enhanced levels of CAT expression from the modified plasmids pHCA47CAT and pHCA177CAT were equivalent and were fivefold higher than that from the modified pHCA0CAT (data not shown). This again indicates that sequences upstream of −47, including the CCAAAT sequence, are not required for the activity of the cardiac actin promoter in L cells. In contrast, sequences at around −110, the region of the CCAAAT sequence, are necessary for the higher-level transcriptional activity seen in C2C12 muscle cells. These sequences may form part of an upstream region of the cardiac actin gene that interacts with muscle-specific, transcription-modulating factors.

Do these upstream regions regulate transcription from a heterologous promoter sequence? The plasmid pA10CAT2 (39), which gives only a low level of CAT activity, can be used to test for the ability of a DNA sequence to regulate the SV40 early promoter (39, 47). This plasmid contains a *Bgl*II cloning site for the insertion of foreign DNA in place of the normal SV40 72-bp regulatory element. Accordingly, we prepared upstream fragments from the human cardiac actin gene and inserted them in this *Bgl*II site. Three fragments were used: (i) a *Pvu*II-*Tth*III fragment (−440 to −47) containing regions D and P; (ii) a *Pvu*II-*Alu*I fragment (−440 to −240) containing region D; (iii) an *Alu*I-*Tth*III fragment (−240 to −47) containing region P. These fragments were inserted next to the SV40 promoter in both orientations (corresponding to sense and antisense in the human cardiac actin gene), and the resulting plasmids were tested for transient CAT expression in C2C12 myotubes as described above.

Insertion of the entire upstream region (D and P) from the cardiac actin gene, in either orientation, led to a four- to fivefold increase in CAT signal in C2C12 myotubes (compare lanes d and e with lanes b and c in Fig. 7). This appears to represent a tissue-specific modulation of the SV40 promoter, because in L cells the D-P fragment did not alter CAT activity (less than twofold compared with that of pA10CAT2; data not shown). However, the resulting CAT signal from pHCA(D-P)SV1CAT in C2C12 myotubes was 20-fold lower than that obtained from the combination of the SV40 72-bp regulatory element with the other SV40 promoter elements (pSV2CAT; Fig. 7, lane a) or the combination of the cardiac actin upstream region with other cardiac actin promoter elements (pHCA485CAT; Fig. 7, lane j). Thus, the action of the cardiac actin upstream sequences is stronger with its own promoter region than with the SV40 heterologous promoter. Whether this is the result of a critical difference in the distance of the regulatory sequences from the promoters in the two constructs (69) or whether it results from a specificity of upstream sequences for particular promoters (17) remains to be evaluated.

Region D or P inserted separately, in either orientation, into the pA10CAT2 plasmid had little effect on CAT activity (Fig. 7, lanes f through i). The modest stimulatory effect of the entire D-P fragment thus may be the result of a sequence element that spans the *Alu*I site at −240, which could be either the 3′ end of region D, if this element is, in fact, several hundred nucleotides long, or a third region that was undetected in the 5′ deletion series experiments. Alternatively, an explanation which would be more consistent with the results from the deletion series would be that the combined effects of two separate sequence elements present in regions D and P are necessary to produce the stimulation of CAT expression.

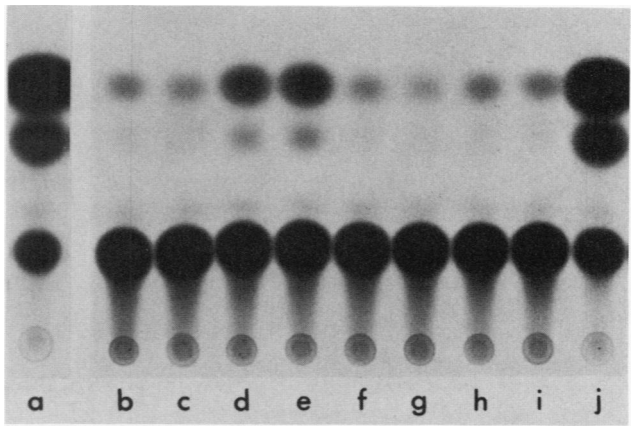

FIG. 7. Regulatory activity of upstream regions of the human cardiac actin gene on the SV40 promoter. DNAs from the plasmid constructions described in the legend to Fig. 6B were transfected into C2C12 cells, and CAT activity was measured as described in the legend to Fig. 3. Lane a, pSV2CAT; lanes b and c, two preparations of pA10CAT2(pSV1CAT); lane d, pHCA(D-P)s.SV1CAT; lane e, pHCA(D-P)a.SV1CAT; lane f, pHCADs.SV1CAT; lane g, pHCADa.SV1CAT; lane h, pHCAPs.SV1CAT; lane i, pHCAPa.SV1CAT; lane j, pHCA485CAT.

DISCUSSION

Evolutionarily conserved upstream sequences in the human cardiac actin gene. The first exon of the human cardiac actin gene is separated from the first coding exon by an intron of approximately 700 bp. Such introns are present in the 5′-untranslated regions but not in the 3′-untranslated regions of all vertebrate actin genes characterized thus far (7). Sequences within these introns might play a role in regulating gene expression (26, 53, 73). Although we cannot exclude this possibility, results of this study and those in the accompanying paper (51) indicate that upstream regions of the human cardiac actin gene, in the absence of introns, are sufficient for tissue-specific transcriptional regulation of this gene. An additional explanation for the presence of an intron in the 5′-untranslated region is that the intron may have been important in the evolution of a new regulatory mode, allowing the newly evolved upstream regulatory region to be reattached to the body of the gene. Such a role for introns in allowing the evolution of new regulatory regions is analogous to the proposed role for introns in the original construction of structural genes by exon shuffling (22).

We have shown previously that in mice and humans a single cardiac actin gene is expressed in both heart and skeletal muscle and is developmentally regulated (28, 49; Minty et al., in press). Results of this study that indicate that the same first exon is present in the cardiac actin mRNAs of these two tissues imply that the developmental regulation of the cardiac actin gene involves differential transcription from a single promoter rather than the differential use of alternative promoters, as has been found for some other muscle

-[A]-

(1)	-197	GCTCCCTATTTGGCCAT	chicken cardiac actin
	-240	GCTCCCTATTTGGCCAT	human cardiac actin
(2)	-165	CTGC-TCCTCACCTGCCTTAGATGGC	chicken cardiac actin
	-215	CTGCCCCCTCCCCTTCCTTACATGGT	human cardiac actin
(3)	-124	CCTTGG--CCATTCATGGCC	chicken cardiac actin
	-157	CCTTGGCTCCATGAATGGCC	human cardiac actin
(4)	-98	GCGCCGAGCCGGCCAAATAAG-AGAAGGTGGC	chicken cardiac actin B
	-93	GCGA--AGCGG--CAAATAAGGAGAAGGTGGC	chicken cardiac actin A
	-119	GCGA--AGGGGACCAAATAAGG-CAAGGTGGC	human cardiac actin
(5)	-34	CTATAAAGCGGC	chicken cardiac actin
	-32	CTATAAAGCGGC	human cardiac actin

-[B]-

(1)	-221	CTCCATATACGGAAAGG	rat skeletal actin
	-177	CTCCTTATACGGAAAGG	chicken skeletal actin
(2)	-168	CCTTCTTTGG	rat skeletal actin
	-128	CCTTCTTTGG	chicken skeletal actin
(3)	-101	GCCCAACACCCAAATATGGC	rat skeletal actin
	-93	GCCCGACACCCAAATATGGC	chicken skeletal actin

-[C]-

(1)	-76	CGTTCCGAAAGTTGCCTTTTATGGC	human β actin
	-74	CGTTCCGAAA-TTGCCTTTTATGGC	rat β actin
	-77	CGCTCCGAAAGTTTCCTTTTATGGC	chicken β actin
(2)	+758	TTGCCTTTTATGGTAAT	human β actin
	+738	TTGCCTTTTATGGTAAT	rat β actin
	+683	TTGCCTTTTATGGTAAT	chicken β actin

-[D]-

(1)	-69	CCAAATTTAGGC	rat α-cardiac MHC
(2)	-80#	CCTATTACAGCCAAAAGTGG	chicken cardiac MLC2
(3)	-122	CCAAGAAAGG	rat skeletal MLC2
(4)	-120	CCAAAATAGC	chicken cardiac troponin T

genes (15). The sequences upstream from the cap site of the human cardiac actin gene show a very high degree of conservation with the corresponding sequences of the chick (8) and mouse (6; I. Garner and M. Buckingham, personal communication) cardiac actin genes, suggesting that these sequences may be involved in the developmental regulation of cardiac actin gene transcription.

Two upstream regions are involved in transcriptional modulation of the cardiac actin gene. In agreement with the suggestion that sequences upstream of the cap site may be involved in developmental regulation of cardiac gene transcription, we found that when the upstream 485 bp of the human cardiac actin gene are linked to the bacterial CAT gene and these plasmids are introduced into cultured cell lines, high levels of CAT expression are obtained in the C2C12 muscle cell line but not in the nonmuscle L or PC-G2 cell lines. Deletion of increasing amounts of the upstream sequence diminishes transcription in muscle cells, suggesting that sequences in this region interact with muscle-specific, positively acting transcriptional regulatory factors (68). Upstream regions likewise have been implicated in the tissue-specific expression of a number of genes, including insulin and chymotrypsin (68), elastase (59), fibroin (66), albumin (60), α-1-antitrypsin (11), β-globin (73), immunoglobulin (26), α-A-crystallin (10), and α-skeletal actin (46) genes; although in several cases it is clear that these are not the only sequences involved.

The use of a progressive series of 5′ deletions showed that at least two regions of upstream sequence from the human cardiac actin gene, D and P, are important in this transcriptional regulation. Region D begins between 443 and 395 bp upstream from the cap site. Deletion to −395 caused an approximately twofold loss in transcriptional activity in muscle cells. Further deletion to −177 had little effect on transcription, at least in our transient assay system. However, deletions to −118, −101, and −47 from the cap site led to a progressive loss in transcriptional activity, and we thus define sequences −177 to −47 as region P. In an attempt to discover potential regulatory elements within these regions, we compared them with known core sequences of viral and cellular regulatory elements (32, 33). In addition, because regulatory mechanisms involved in muscle differentiation appear to be evolutionarily conserved (4, 56), we looked for conserved sequences in the upstream regions of the human, mouse, and chick cardiac actin genes. Finally, we examined the upstream sequences of other contractile protein genes expressed in striated muscle.

Region D contains sequences homologous to enhancer core sequences, but these sequences do not, on their own, have enhancer potential. Enhancer sequences in cellular and viral genes have been defined by their ability to stimulate transcription from heterologous promoters, in either orientation, over distances of several kilobases (36). An enhancer core sequence was originally identified by Weihler et al. (70) and has recently been implicated in the rescue of defective SV40 virus mutants (34). A sequence at −390 in the human cardiac actin gene, TGTGGCTTATTGTCCCCAAGGCT, contains several regions (underlined) which are also present in the core sequence of the SV40 enhancer. In fact, these human cardiac actin sequences match the SV40 sequence better than do other viral and cellular enhancers, even that of the closely related polyomavirus (32, 33).

The mouse cardiac actin gene contains the sequence $(TG)_{24}$ in place of the TGTGG residues at −390 in the human gene (6; I. Garner, personal communication). Such a sequence constitutes a potential Z-DNA sequence and has been shown to have enhancer potential (31). This sequence is absent from the human gene in this region, although, as described by Hamada et al. (30), a TG repeat is present in the human cardiac actin gene in the fifth intron. Within region D, however, there is another potential Z-DNA region, a stretch of eight alternating purines and pyrimidines (TGCATGTG), in front of and including the TGTGG sequence mentioned above. This is a length sufficient to form Z-DNA (54) and to contribute to enhancer function (34) in SV40.

Region D did not have enhancer potential when we tested it linked to the heterologous SV40 promoter in plasmid pA10CAT2 (39, 47), although this region did have an effect in combination with region P. It would thus seem that if these two sequence elements (the corelike sequence and the putative Z-DNA-forming sequence) contribute to the regulatory activity of region D, they must do so by influencing the activity of a second regulatory sequence rather than acting as part of an autonomously functioning enhancer element.

Region P contains a repeated motif $CC(A/T)_6GG$ that is conserved in evolution. There is a striking degree of sequence conservation (~70%) between the region −240 to −40 of the human cardiac actin gene and −200 to −35 of the chick gene (8) (Fig. 2). Despite this similarity, the human gene does not

FIG. 8. (A) Evolutionarily conserved sequences in the upstream regions of human and chick cardiac actin genes. The 240 bp upstream of the cap sites of the human and chicken cardiac actin genes were compared (Fig. 2). Regions of similarity containing subregions showing 9 of 10 or 10 of 12 matched bases are shown (asterisks). The second set of numbers indicates nucleotides from the cap site, and the CCArGG motif is underlined in each case. The chicken sequence is taken from Chang et al. (8). While this study was in preparation, Eldridge et al. (19) also published a nucleotide sequence for a chicken cardiac actin gene which showed some nucleotide differences from the sequence described by Chang et al. (8) in this region. Most of these changes did not alter the overall homology with either the human gene or the CCArGG sequences. However, in one case, for region number 4, the two chicken sequences are substantially different, and we indicate both sequences. A is from Chang et al. (8) and B is from Eldridge et al. (19). (B) Presence of CCArGG sequences in skeletal muscle actin genes. The upstream sequences of the rat and chicken skeletal actin genes were from Ordahl and Cooper (58) and Nudel et al. (56). Region 1 in the rat gene was only partially homologous with the chicken sequence at −238 (one mismatch and two gaps), as aligned by Nudel et al. (56). This alignment involves the introduction of a 55-bp gap in the chicken sequence. However, the chicken sequence at −176 showed a greater similarity (only one mismatch) with this region of the rat gene, and this alignment required only a 5-bp gap in the chicken sequence. (C) Presence of CCArGG sequences in β-cytoplasmic actin genes. Rat, chick, and human β-cytoplasmic actin gene sequences were from Nudel et al. (57), Kost et al. (38), and Ng et al. (53). CCArGG sequences are underlined. (D) Presence of CCArGG-like sequences in other contractile protein genes. The regions between 50 and 150 bp upstream of the cap sites of other contractile protein genes were searched for the presence of sequences resembling the CCArGG sequences found in vertebrate actin genes. MHC, Myosin heavy chain; MLC, myosin light chain. The sequences were from Mahdavi et al. (41) (region 1), Winter et al. (71) (region 2), Nudel et al. (55) (region 3), and Cooper and Ordahl (12) (region 4). Sequences corresponding exactly to our definition of the CCArGG motif (see text) are underlined. Numbers indicate base pairs upstream from the cap site. For the chicken cardiac myosin light chain 2, the cap site is not known, and this number is an approximate estimate based on the fact that this sequence is 51 bp upstream from the TATA sequence.

contain the sequences CCGGGCCCGACA or GCCGCGC CGG which have been found to be common to the chick cardiac and skeletal actin genes (8). Nor does it contain the hexanucleotide GGGCGG or CCGCCC, which represent binding sites for the mammalian transcription factor Sp1 (17). However, five short sequence blocks in the chick and human genes are highly conserved and contain shorter elements that are identical at 9 of 10 or 10 of 12 nucleotides. These are listed as segments 1 through 5 in Fig. 8A.

The perfect conservation of the 17-nucleotide segment 1 (Fig. 8A) between the human and chick genes, which are separated by 250 million years of evolution, strongly suggests a functional role for this sequence. However, deletion of this sequence has little effect on transcriptional activity, as measured in our transient CAT assay. This might be explained if there were redundant copies of regulatory signals contained within region P. For example, in the *Drosophila melanogaster hsp-70* gene (62) and in the mouse immunoglobulin VH gene (43), multiple consensus regulatory sequences are present in the 5'-flanking region of the genes, but only the TATA proximal copy is essential for expression of genes transfected into cultured cells. In the case of the *hsp-70* gene it has been shown that both copies contribute to maximal expression in vivo in *D. melanogaster* (62).

While there are no strong similarities between the short sequence blocks 1 to 4 (Fig. 8A) that are conserved between the human and chick cardiac actin genes, these regions do share a common motif, $CC(A/T)_6GG$ (designated CC--A-rich--GG), in which up to one GC base pair is allowed in the AT-rich center. These CCArGG sequences are underlined in Fig. 8A. The CCArGG sequence in region 1 and in region 4 (the CCAAAT sequence; see below) do show a striking similarity when the complement of one of the sequences is considered (e.g., region 1 [GGCCAAATAGGGAG]). This observation supports the possibility that the different CCArGG sequences represent a family of regulatory elements and suggests that these elements may be bidirectional. CAT plasmids containing 177, 118, and 47 nucleotides of upstream sequence from the human cardiac actin gene (which thus contain two, one, or none of these four CCArGG sequences, respectively) show a progressive loss of CAT signal. Deletion from −118 to −101, thereby eliminating the first eight nucleotides of the last CCArGG sequence, causes a substantial, though not complete, loss of CAT activity. Thus, these CCArGG sequences seem to be prime candidates for an important part of a series of small regulatory elements, based on their evolutionary conservation and their occurrence in regions defined as having regulatory activity in our transient CAT assay system.

A similar CCArGG motif is found in other contractile protein genes. Sequence comparisons of the rat and chick skeletal muscle actin genes also show several small blocks of conserved sequence in the upstream 220 bp (56, 58). These conserved segments contain CCArGG sequences (underlined in Fig. 8B). Similarly, a conserved CCArGG sequence, TTGCCTTTTATGG, is found between the CCAAT and TATA boxes of the chick, rat, and human cytoplasmic β-actin genes (38, 53, 57) (Fig. 8C, region 1), and the human γ-actin gene also contains a CCArGG-like sequence, GC CATATATGG, in this 5' upstream region (H. Erba, unpublished data). The sequence TTGCCTTTTATGG is also found in the first introns of the human, rat, and chick β-actin genes (53) (Fig. 8C, region 2). As expected, a short and relatively nonspecific consensus sequence such as the CCArGG sequence occurs within and around a large number of other genes. However, the evolutionary conservation of these sequences and their locations in the vicinity of vertebrate actin gene promoters suggest that they may play a role in the transcriptional regulation of these genes. The presence of CCArGG sequences in potential regulatory regions of the two nonmuscle actin genes suggests that the presence of such sequences is not by itself sufficient to account for the tissue-specific expression of these members of the actin gene family. Several other contractile protein genes expressed in striated muscle contain a CCArGG or CCArGG-like region within 50 to 150 nucleotides upstream of the cap site (Fig. 8D), although this is not true for all such genes (e.g., the alkali light chains [15]).

CCArGG region 4 (Fig. 8A) has been designated as the CAAT box (2, 18) of the chick cardiac actin gene (8), and CCArGG region 3 (Fig. 8B) has been designated as the CAAT box of the rat and chick skeletal muscle actin genes (58). However, these sequences match poorly with the original consensus CAAT sequence (GGC/TCAATCT) described by Benoist et al. (2) and is further upstream (−107) in the human cardiac actin gene than the sequences compared previously (2) (around −60 to −70). Similarly, the CCArGG sequences found in the other muscle genes (Fig. 8D) do not necessarily correspond to a consensus CAAT sequence. A role for the CAAT box as part of a basal promoter sequence stems from work on β-globin genes by Dierks et al. (16) and Charnay et al. (9). As discussed by Dierks et al. (16) and Charnay et al. (9), the previous results cannot necessarily be generalized to other genes. In addition, recent work on a naturally occurring single-base mutation in the CCAAT box region of a fetal globin gene indicates that this sequence may modulate developmental, as well as basal, expression (21). The observations made in this study, as well as those in the accompanying paper (51), indicate that the CCAAAT sequence in the human cardiac actin gene plays a role in tissue-specific transcription rather than as part of a basal promoter sequence.

Although the precise sequences of adenosines and thymines of the AT-rich regions of the different CCArGG elements are different, the sequence of a particular CCArGG element is often conserved in evolution (Fig. 8). This is especially striking for the β-actin gene. Thus, the precise sequence of these regions indeed may be important, for example, in dictating affinities of regulatory proteins for these DNA regions. In addition, the CCArGG elements are each imbedded in larger segments that are conserved in evolution but are not related to each other. These larger elements may contribute to the specificity of interaction with regulatory factors. We suggest that the larger elements containing the CCArGG sequences in the cardiac actin genes may have evolved to interact with particular sets of transcription-modulating factors (17, 35) that are responsible for transcriptional regulation during muscle development.

ACKNOWLEDGMENTS

We are grateful to Edna Hardeman and Peter Gunning for gifts of RNA samples, Helen Blau for C2C12 muscle cells, Chantal Rabourdin-Combe for Ltk⁻ cells, Bruce Howard and L. Laimins for plasmids pSV2CAT and pA10CAT2, and J. Garner and M. Buckingham for communicating sequence data on the mouse cardiac actin gene. We also thank Peter Evans and Patricia Benton-Vosman for invaluable help with plasmid DNA preparations and Peter Gunning and Helen Blau for many helpful discussions and thoughtful readings of the manuscript.

A.M. was a fellow of the Fogarty International Center. This work was supported in part by grants from the Muscular Dystrophy

Association, the National Institutes of Health, and the Veterans Administration to L.K.

LITERATURE CITED

1. **Bains, W., P. Ponte, H. Blau, and L. Kedes.** 1984. Cardiac actin is the major actin gene product in skeletal muscle cell differentiation in vitro. Mol. Cell. Biol. **4**:1449–1453.
2. **Benoist, C., K. O'Hare, R. Breathnach, and P. Chambon.** 1980. The ovalbumin gene-sequence of putative control regions. Nucleic Acids Res. **8**:127–142.
3. **Bird, A., M. Taggart, M. Frommer, O. J. Miller, and D. Macleod.** 1985. A fraction of the mouse genome that is derived from islands of nonmethylated, CpG-rich DNA. Cell **40**:91–99.
4. **Blau, H. M., C.-P. Chiu, and C. Webster.** 1983. Cytoplasmic activation of human nuclear genes in stable heterokaryons. Cell **32**:1171–1180.
5. **Borelli, E., R. Hen, and P. Chambon.** 1984. Adenovirus-2 E1A products repress enhancer-induced stimulation of transcription. Nature (London) **312**:608–612.
6. **Buckingham, M. E., S. Alonso, P. Barton, G. Bugaisky, A. Cohen, P. Daubas, I. Garner, A. Minty, B. Robert, and A. Weydert.** 1984. Molecular biology of development: actin and myosin genes and their expression during myogenesis in the mouse. Alan R. Liss, Inc., New York.
7. **Buckingham, M. E., and A. J. Minty.** 1983. Eucaryotic genes: their structure, activity and regulation; contractile protein genes, p. 365–395. Butterworths, London.
8. **Chang, K. S., K. N. Rothblum, and R. J. Schwartz.** 1985. The complete sequence of the chicken α-cardiac actin gene: a highly conserved vertebrate gene. Nucleic Acids Res. **13**:1223–1237.
9. **Charnay, P., P. Mellon, and T. Maniatis.** 1985. Linker scanning mutagenesis of the 5'-flanking region of the mouse β-major-globin gene: sequence requirements for transcription in erythroid and nonerythroid cells. Mol. Cell. Biol. **5**:1498–1511.
10. **Chepelinsky, A. B., C. R. King, P. S. Zelenka, and J. Piatigorsky.** 1985. Lens-specific expression of the chloramphenicol acetyltransferase gene promoted by 5' flanking sequences of the murine α-A-crystallin gene in explanted chicken lens epithelia. Proc. Natl. Acad. Sci. USA **82**:2334–2338.
11. **Ciliberto, G., L. Dente, and R. Cortese.** 1985. Cell-specific expression of a transfected human α1-antitrypsin gene. Cell **41**:531–540.
12. **Cooper, T. A., and C. P. Ordahl.** 1985. A single cardiac troponin T gene generates embryonic and adult isoforms via developmentally regulated alternate splicing. J. Biol. Chem. **260**: 11140–11148.
13. **Czosnek, H., U. Nudel, Y. Mayer, P. E. Barker, D. D. Pravtcheva, F. R. Ruddle, and D. Yaffe.** 1983. The genes coding for the cardiac muscle actin, the skeletal muscle actin and the cytoplasmic β-actin are located on three different mouse chromosomes. EMBO J. **2**:1977–1979.
14. **Czosnek, H., U. Nudel, M. Shani, P. E. Barker, D. D. Pravtcheva, F. Ruddle, and D. Yaffe.** 1982. The genes coding for the muscle contractile proteins, myosin heavy chain, myosin light chain 2, and skeletal muscle actin are located on three different mouse chromosomes. EMBO J. **1**:1299–1305.
15. **Daubas, P., B. Robert, I. Garner, and M. Buckingham.** 1985. A comparison between mammalian and avian fast skeletal muscle alkali light chain genes: regulatory implications. Nucleic Acids Res. **13**:4623–4643.
16. **Dierks, P., A. van Oyen, M. D. Cochran, C. Dobkin, J. Reiser, and C. Weissman.** 1983. Three regions upstream from the cap site are required for efficient and accurate transcription of the rabbit β-globin gene in mouse 3T6 cells. Cell **32**:695–706.
17. **Dynan, W. S., and R. Tijan.** 1985. Control of eucaryotic messenger RNA synthesis by sequence-specific DNA-binding proteins. Nature (London) **316**:774–778.
18. **Efstratiadis, A., J. W. Posakony, T. Maniatis, R. Lawn, C. O'Connell, R. A. Spritz, J. K. DeRiel, B. G. Forget, S. M. Weissman, J. L. Slightom, A. E. Blechl, O. Smithies, F. E. Baralle, C. C. Shoulders, and N. J. Proudfoot.** 1980. The structure and the evolution of the human β-globin gene family. Cell **21**:653–668.
19. **Eldridge, J., Z. Zehner, and B. M. Paterson.** 1985. Nucleotide sequence of the chicken cardiac α actin gene: absence of strong homologies in the promoter and 3'-untranslated regions with the skeletal α actin sequence. Gene **36**:55–63.
20. **Engel, J., P. Gunning, and L. Kedes.** 1982. Human cytoplasmic actin proteins are encoded by a multigene family. Mol. Cell. Biol. **2**:674–684.
21. **Gelinas, R., B. Endlich, C. Pfeiffer, M. Yagi, and G. Stamatoyannopoulos.** 1985. G to A substitution in the distal CCAAT box of the $^{A}\gamma$-globin gene in Greek hereditary persistence of fetal haemoglobin. Nature (London) **313**:323–325.
22. **Gilbert, W.** 1978. Why genes in pieces? Nature (London) **271**:501.
23. **Goodman, R., and H. R. Herschman.** 1978. Nerve growth factor-mediated induction of tyrosine hydroxylase in a clonal pheochromocytoma cell line. Proc. Natl. Acad. Sci. USA **75**:4587–4590.
24. **Gorman, C. M., L. F. Moffat, and B. H. Howard.** 1982. Recombinant genomes which express chloramphenicol acetyltransferase in mammalian cells. Mol. Cell. Biol. **2**:1044–1051.
25. **Graham, F. L., and A. J. van der Eb.** 1973. A new technique for the assay of infectivity of human adenovirus 5 DNA. Virology **52**:456–467.
26. **Grosschedl, R., and D. Baltimore.** 1985. Cell-type specificity of immunoglobulin gene expression is regulated by at least three DNA sequence elements. Cell **41**:885–897.
27. **Gunning, P., T. Mohun, S.-Y. Ng, P. Ponte, and L. Kedes.** 1984. Evolution of the human sarcomeric-actin genes: evidence for units of selection within the 3' untranslated regions of the mRNAs. J. Mol. Evol. **20**:202–214.
28. **Gunning, P., P. Ponte, H. Blau, and L. Kedes.** 1983. α-Skeletal and α-cardiac actin genes are coexpressed in adult human skeletal muscle and heart. Mol. Cell. Biol. **3**:1985–1995.
29. **Gunning, P., P. Ponte, L. Kedes, R. Eddy, and T. Shows.** 1984. Chromosomal location of the co-expressed human skeletal and cardiac actin genes. Proc. Natl. Acad. Sci. USA **81**:1813–1817.
30. **Hamada, H., M. G. Petrino, and T. Kakunaga.** 1982. Molecular structure and evolutionary origin of human cardiac muscle actin gene. Proc. Natl. Acad. Sci. USA **79**:5901–5905.
31. **Hamada, H., M. Seidman, B. Howard, and C. M. Gorman.** 1984. Enhanced gene expression by the poly(dT-dG) · poly(dC-dA) sequence. Mol. Cell. Biol. **4**:2622–2630.
32. **Hen, R., E. Borelli, P. Sassone-Corsi, and P. Chambon.** 1983. An enhancer element is located 340 base pairs upstream from the adenovirus-2 E1A capsite. Nucleic Acids Res. **11**:8747–8760.
33. **Hermobel, P., B. Bourachot, and M. Yaniv.** 1984. Two distinct enhancers with different cell specificities coexist in the regulatory region of polyoma. Cell **39**:653–662.
34. **Herr, W., and Y. Gluzman.** 1985. Duplications of a mutated SV40 enhancer restore its activity. Nature (London) **313**: 711–714.
35. **Jones, K. A., K. R. Yamamoto, and R. Tijan.** 1985. Two distinct transcription factors bind to the HSV thymidine kinase promoter in vitro. Cell **39**:559–572.
36. **Khoury, G., and P. Gruss.** 1983. Enhancer elements. Cell **33**:313–314.
37. **Killary, A. M., and R. E. K. Fournier.** 1984. A genetic analysis of extinction: trans-dominant loci regulate expression of liver-specific traits in hepatoma hybrid cells. Cell **38**:523–534.
38. **Kost, T. A., N. Theodorakis, and S. H. Hughes.** 1983. The nucleotide sequence of the chick cytoplasmic actin gene. Nucleic Acids Res. **11**:8287–8301.
39. **Laimins, L. A., G. Khoury, C. Gorman, B. Howard, and P. Gruss.** 1982. Host specific activation of transcription by tandem repeats from simian virus 40 and Moloney sarcoma virus. Proc. Natl. Acad. Sci. USA **79**:6453–6457.
40. **Lopata, M., D. W. Cleveland, and B. Sollner-Webb.** 1984. High-level transient expression of a chloramphenicol acetyltransferase gene by DEAE dextran mediated DNA transfection coupled with a dimethylsulfoxide or glycerol shock treatment. Nucleic Acids Res. **12**:5707–5717.
41. **Mahdavi, V., A. P. Chambers, and B. Nadal-Ginard.** 1984.

Cardiac α- and β-myosin heavy chains are organized in tandem. Proc. Natl. Acad. Sci. USA **81**:2626–2630.
42. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
43. **Mason, J. O., G. T. Williams, and M. S. Neuberger.** 1985. Transcription cell type specificity is conferred by an immunoglobulin V_H promoter that includes a functional consensus sequence. Cell **41**:479–487.
44. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific cleavage reactions. Methods Enzymol. **65**:499–560.
45. **Mayer, Y., H. Czosnek, P. E. Zeelon, D. Yaffe, and U. Nudel.** 1984. Expression of the genes coding for the skeletal muscle and cardiac actins in the heart. Nucleic Acids Res. **12**:1087–1100.
46. **Melloul, D. D., B. Aloni, J. Calvo, D. Yaffe, and U. Nudel.** 1984. Developmentally regulated expression of chimeric genes containing muscle actin DNA sequences in transfected myogenic cells. EMBO J. **3**:983–990.
47. **Mercola, M., J. Goverman, C. Mirell, and K. Calame.** 1985. Immunoglobulin heavy-chain enhancer requires one or more tissue-specific factors. Science **227**:266–270.
48. **Messing, J.** 1983. New M13 vectors for cloning. Methods Enzymol. **101**:28–78.
49. **Minty, A. J., S. Alonso, M. Caravatti, and M. E. Buckingham.** 1982. A fetal skeletal muscle actin mRNA in the mouse and its identity with cardiac actin mRNA. Cell **30**:185–192.
50. **Minty, A. J., S. Alonso, J. L. Guenet, and M. Buckingham.** 1983. Number and organization of actin-related sequences in the mouse genome. J. Mol. Biol. **167**:77–101.
51. **Minty, A., H. Blau, and L. Kedes.** 1986. Two-level regulation of cardiac actin gene transcription: muscle-specific modulating factors can accumulate before gene activation. Mol. Cell. Biol. **6**:2146–2157.
52. **Mohun, T. J., S. Brennan, N. Dathan, S. Fairman, and J. B. Gurdon.** 1984. Cell type specific activation of actin genes in the early amphibian embryo. Nature (London) **311**:716–721.
53. **Ng, S.-Y., P. Gunning, R. Eddy, P. Ponte, J. Leavitt, T. Shows, and L. Kedes.** 1985. Evolution of the functional human β-actin gene and its multi-pseudogene family: conservation of noncoding regions and chromosomal dispersion of pseudogenes. Mol. Cell. Biol. **5**:2720–2732.
54. **Nordheim, A., and A. Rich.** 1983. Negatively supercoiled SV40 DNA contains Z-DNA segments within transcriptional enhancer sequences. Nature (London) **303**:674–679.
55. **Nudel, U., J. Calvo, M. Shani, and Z. Levy.** 1984. The nucleotide sequence of a rat myosin light chain 2 gene. Nucleic Acids Res. **12**:7175–7186.
56. **Nudel, U., D. Greenberg, C. P. Ordahl, O. Saxel, S. Neuman, and D. Yaffe.** 1985. Developmentally regulated expression of a chicken muscle-specific gene in stably transfected rat myogenic cells. Proc. Natl. Acad. Sci. USA **82**:3106–3109.
57. **Nudel, U., R. Zakut, M. Shani, S. Neuman, Z. Levy, and D. Yaffe.** 1983. The nucleotide sequence of the rat cytoplasmic β-actin gene. Nucleic Acids Res. **11**:1759–1771.
58. **Ordahl, C. P., and T. A. Cooper.** 1983. Strong homology in promoter and 3′-untranslated regions of chick and rat alpha-actin genes. Nature (London) **303**:348–349.
59. **Ornitz, D. M., R. D. Palmiter, R. E. Hammer, R. L. Brinster, G. H. Swift, and R. J. MacDonald.** 1985. Specific expression of an elastase-human growth hormone fusion gene in pancreatic acinar cells of transgenic mice. Nature (London) **313**:600–602.
60. **Ott, M. O., L. Sperling, P. Hermobel, M. Yaniv, and M. C. Weiss.** 1984. Tissue-specific expression is confered by a sequence from the 5′ end of the rat albumin gene. EMBO J. **3**:2505–2510.
61. **Paterson, B. M., and J. D. Eldridge.** 1984. α-Cardiac actin is the major sarcomeric isoform expressed in embryonic avian skeletal muscle. Science **224**:1436–1438.
62. **Pelham, H.** 1985. Activation of heat shock genes in eukaryotes. Trends Genet. **1**:31–35.
63. **Robert, B., P. Barton, A. Minty, P. Daubas, A. Weydert, F. Bonhomme, J. Catalan, D. Chazottes, J.-L. Guenet, and M. Buckingham.** 1985. Use of an inter-specific mouse back-cross to investigate genetic linkage between myosin and actin genes. Nature (London) **314**:181–183.
64. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
65. **Seiler-Tuyns, A., J. D. Eldridge, and B. M. Paterson.** 1984. Expression and regulation of chicken actin genes introduced into mouse myogenic and nonmyogenic cells. Proc. Natl. Acad. Sci. USA **81**:2980–2984.
66. **Tsuda, M., and Y. Suzuki.** 1983. Transcription modulation in vitro of the fibroin gene exerted by a 200 base-pair region upstream from the TATA box. Proc. Natl. Acad. Sci. USA **80**:7442–7446.
67. **Vandekerckhove, J., and K. Weber.** 1979. The complete amino acid sequence of actins from bovine aorta, bovine heart, bovine fast skeletal muscle, and rabbit slow skeletal muscle. Differentiation **14**:123–133.
68. **Walker, M. D., T. Edlund, A. M. Boulet, and W. J. Rutter.** 1983. Cell-specific expression controlled by the 5′-flanking region of insulin and chymotrypsin genes. Nature (London) **306**:557–561.
69. **Wasylyk, B., C. Wasylyk, and P. Chambon.** 1984. Short and long range activation by the SV40 enhancer. Nucleic Acids Res. **12**:5589–5608.
70. **Weihler, H., M. Konig, and P. Gruss.** 1983. Multiple point mutations affecting the SV40 virus enhancer. Science **219**: 626–631.
71. **Winter, B., H. Klapthor, K. Wiebauer, H. Delius, and H. Arnold.** 1985. Isolation and characterization of the chicken cardiac myosin light chain (L-2A) gene. J. Biol. Chem. **260**:4478–4483.
72. **Wolf, S. F., and B. R. Migeon.** 1985. Clusters of CpG dinucleotides implicated by nuclease hypersensitivity as control elements of housekeeping genes. Nature (London) **314**:467–469.
73. **Wright, S., A. Rosenthal, R. Flavell, and F. Grosveld.** 1984. DNA sequences required for regulated expression of β-globin genes in murine erythroleukemia cells. Cell **38**:265–273.
74. **Yaffe, D., and O. Saxel.** 1977. Serial passaging and differentiation of myogenic cells isolated from dystrophic mouse muscle. Nature (London) **270**:725–727.